



School of Law

# **BENJAMIN KLEIN**

**Joshua D. Wright,  
George Mason University School of Law**

***Pioneers of Law and Economics* (Edward Elgar  
Publishing, Lloyd R. Cohen and Joshua D. Wright,  
eds.), Forthcoming**

**George Mason University Law and Economics  
Research Paper Series**

**08-31**

This paper can be downloaded without charge from the Social Science  
Research Network at [http://ssrn.com/abstract\\_id=1143568](http://ssrn.com/abstract_id=1143568)

# BENJAMIN KLEIN

Joshua D. Wright<sup>^</sup>

Forthcoming in PIONEERS OF LAW AND ECONOMICS  
Lloyd R. Cohen and Joshua D. Wright, eds., Elgar Publishing, 2008

## **Abstract:**

This chapter in the book PIONEERS OF LAW AND ECONOMICS explores the contributions of Benjamin Klein to law and economics. I explore the intellectual foundations of Klein's pioneering analysis of the hold-up problem, the theory of the firm, vertical restraints, franchising, and the role of contract terms in facilitating self-enforcement of contractual relationships. I also discuss the significant influence of Klein's work on antitrust law, as well as its implications for contract interpretation. Klein's pioneering work over the past 30 years has not only provided us with a much greater understanding of contractual arrangements, but also a model for law and economics scholars and economists interested in explaining real world phenomenon rather than merely producing blackboard insights.

---

<sup>^</sup> Visiting Professor, University of Texas School of Law; Assistant Professor, George Mason University School of Law (on leave). I thank Benjamin Klein and Scott Masten for valuable comments and suggestions. Brandy Wagstaff provided excellent research assistance.

## I. INTRODUCTION

Benjamin Klein's influential contributions to industrial organization, the theory of the firm, and the economics of contracts are widely recognized by economists. A recent analysis of articles published in peer-reviewed economics journals since 1970 reveals two of Klein's publications among the top 60 cited. Klein, along with Oliver Williamson, is also credited with explaining the economic relationship between asset specificity and vertical integration that has been described as "one of the great success stories in industrial organization over the last 25 years."

While Klein's contributions to the economics of contracts and to the theory of the firm are widely recognized in the economics literature, Klein's work has also been highly influential in the economic analysis of the law. This essay explores the intellectual foundations of these contributions, their impact on law and economics, and considers some potential applications of Klein's insights to contract interpretation.

It is customary in essays of this genre to define the scope of the project and issue some disclaimers about obvious omissions. The primary purpose of this essay is to identify and explore the intellectual foundations and themes in Klein's pioneering work in law and economics. In order to economize on page constraints, I will focus almost exclusively on Klein's contributions to the law and economics of contracts. "Contracts" is defined broadly in this setting to include implicit and self-enforced agreements, specific contract terms, and particular contractual arrangements such as franchising and vertical integration.

This organizational framework omits some important work that would obviously merit inclusion in an essay cataloging Klein's contributions to law and economics. If the weight of an intellectual contribution was properly weighted

by citation or some other metric of scholarly impact, one might reasonably start a chapter such as this with Klein's seminal analysis of litigation and settlement with George Priest, *The Selection of Disputes for Litigation*. Priest and Klein (1984) has proven to be one of the most influential articles in legal scholarship, and is certainly part of the law and economics canon.<sup>1</sup> The Priest-Klein litigation model explored the systematic differences between litigated and settled cases and generated, amongst other insights, the well-known and much tested prediction that plaintiffs will prevail at trial 50 percent of the time under certain conditions, regardless of the likelihood with which they would have won the underlying cases they settled. Subsequent researchers have built upon, modified, extended, and tested the Priest-Klein model in various ways.<sup>2</sup> Despite the Priest-Klein model's obvious importance in the development of an economic analysis of settlement and litigation, it will not be covered in detail in this essay.

A second important omission of a pathbreaking analysis is Granitz & Klein's (1996) account of the famous Standard Oil monopoly. Contrary to the conventional wisdom in McGee's (1958) classic article that Standard Oil owed its success to serial mergers and acquisitions executed at or above market prices, Granitz & Klein convincingly document that Standard Oil's monopolization strategy involved Rockefeller's cartelization of the railroad industry by serving as the "cartel ringmaster." In other words, Standard Oil used its dominance in refining to effectively police the railroads in order to stabilize the petroleum cartel. Granitz & Klein's analysis of Standard Oil continues to be the one of the only convincing empirical examples in the literature of the well-known "Raising

---

<sup>1</sup> See, e.g., Shapiro (1996) (ranked 99<sup>th</sup> out of all law articles); Krier & Schwab (1997) (ranked 81<sup>st</sup> of all law articles); Landes & Posner (YEAR) (ranked 28<sup>th</sup> of all law articles).

<sup>2</sup> Empirical work includes Priest (1987); Eisenberg (1990); Waldfogel (1995); Kessler, Meites, and Miller (1996); Siegelman and Donohue (1995); and Froeb (1993). Theoretical contributions include Grossman and Katz (1983); Hylton (1993); Shavell (1995); Spier (1992); and Nalebuff (1987).

Rivals' Costs" phenomenon. As Klein (2003) points out in relationship to theories of exclusion, the relative absence of empirical proof combined with the fact that the Standard Oil example involved a *horizontal* conspiracy suggests that antitrust policy that expends significant resources attempting to identify potentially anticompetitive vertical agreements is not yet empirically justified. As with Priest and Klein (1984), and with apologies to the interested reader, a more detailed discussion is omitted.

One additional task remains before we turn to Klein's scholarly contributions to the law and economics of contracts. Contributions to law and economics can come in many forms other than scholarship. I mention some of Klein's important contributions of this variety before moving on to the heart of the essay. Klein currently serves on the board of editors of several academic publications focusing on economic analysis of law, including *Supreme Court Economic Review*, *Antitrust Law Journal*, and the *Journal of Law, Economics & Organization*. Klein is also on the Advisory Board of the *New Palgrave Dictionary of Economics and Law*. Klein has also served as a consultant to the Federal Trade Commission (FTC) and the Department of Justice Antitrust Division and taught at the Economics Institute for Federal Judges. Klein also served as the Economics Director for the once active Joint Degree Program in Law and Economics at UCLA. Klein has also had considerable influence over subsequent generations of law and economics scholars in law and economics who were former students at UCLA, including the author. Timothy Muris (2003), a former student and the former Chairman of the FTC, singles out Klein's contributions to law and economics as having substantial influence on the development of modern competition policy.

I would also be remiss if I did not take this opportunity to acknowledge and express gratitude for Klein's contribution to my own academic career as an

economics graduate student at UCLA, a research assistant, and later as a co-author and mentor. Perhaps the most important lesson I learned was that economics was not about theoretical blackboard exercises, mathematical models, or regressions, but using all of the available tools to explain real world phenomena. Economics is at its best when it used to shed light on and increase our understanding of actions, relationships, and institutions that are not well understood. There is an especially significant cost to theoretical abstraction in the world of law and economics because institutions, the content of the law, and how courts enforce the law affect the incentives of the firms and individuals and economic agents being studied. I was taught that economic analysis is a process that requires “getting one’s hands dirty” with the facts and viewing and interpreting those facts through the lens of economic theory. While I consider my self fortunate to have received these lessons in the classroom, they are evident in Klein’s approach to economics, and, as we shall see, apparent in his body of work.

Klein’s approach to economics, in turn, no doubt began with his economics training at the University of Chicago, under the tutelage of two of the primary intellectual influences in his career: Milton Friedman, his dissertation advisor, and George Stigler. As one might predict from a student of Friedman and Stigler, Klein’s two primary fields were monetary theory and industrial organization. Klein’s dissertation, “*The Payment of Interest in Commercial Bank Deposits and The Price of Money: A Study of Demand for Money,*” reflected Klein’s interest in each of these fields and would later be published in the *American Economic Review* as “*Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money.*” Klein’s analysis corrected previous attempts to estimate the demand for money by estimating and allowing for perfectly competitive interest payments on commercial bank deposits, allowing own price and a cross

price to have separate influences, and assuming that the prohibition of interest payments is totally ineffective.

Little in the topic of that particular publication would allow the casual observer to confidently predict Klein's subsequent contributions to industrial organization, contracts, antitrust, or law and economics. Nonetheless, the dissertation is notable for two reasons. First, it identifies Friedman and Stigler as two significant intellectual influences in Klein's work. Second, Klein's work in monetary theory is relevant to this essay because it provides the seeds for what would ultimately produce his major contributions to law and economics. Specifically, the origins of Klein's framework for analyzing contractual relationships can be found in the dissertation's second chapter, later published in the *Journal of Money, Credit, and Banking* as "*The Competitive Supply of Money.*"

In that paper, Klein explored the widely accepted notion that government intervention in the money industry was justified. What were the money industry's unique characteristics that required the monopolistic supply of currency? Klein argued that a system of competing money suppliers would not necessarily generate a hyper-inflation and an infinitely high level of money prices so long as the firms were able to prevent counterfeiting. Where the private supply of money raises the opportunity for the firms to "deceive" customers by supplying more money than anticipated ex ante, firms can rely on their brand name as a capital asset that facilitates performance. Klein's work over the next several years would be largely devoted to building upon that framework, which would be incomplete without the combination of two fundamental economic insights involving the role of contract terms in facilitating contractual performance.

The first key insight comes from Klein's work with Armen Alchian and Robert Crawford ("KCA") on asset specificity and the economic forces creating

the well-known “hold-up” problem. While often described as an analysis of the existence of or conditions that give rise to the hold-up problem, a more precise description is that it was an attempt to provide a framework for understanding the contractual responses of transactors anticipating potential hold-up problems. KCA explain the use of vertical integration as an efficient contractual response to potential hold-up problems associated with asset-specific investments. While this analysis is properly viewed as a seminal contribution to the theory of the firm, it is important to also recognize it as the first step in the development of a broader analytical framework for understanding the role of contract terms in facilitating contractual performance.

Klein and Leffler (1981) extend this framework by analyzing private contract enforcement mechanisms. Because private enforcement capital is limited, and written contracts are necessarily imperfect and incomplete, transactors must combine both court-enforced written terms with self-enforced unwritten terms to define what Klein calls the contractual relationship’s “self-enforcing range.” Written contract terms commit the parties with respect to certain actions that might effectuate a hold-up at the cost of creating contractual rigidity. In many cases, transactors may elect to avoid the costs associated with court-enforcement of written contract terms by intentionally leaving elements of performance unspecified and enforced through private enforcement mechanisms.

Together, these two economic insights would be used to analyze the role of contract terms in facilitating self-enforcement and become a fertile basis for increasing our knowledge about contracts and the contracting process when applied to specific settings. For example, the self-enforcement framework would provide the basis for contributions to the theory of the firm, the economics of vertical restraints, franchises, block booking, exclusive dealing, slotting contracts,

and various “non--standard” contract terms in facilitating contractual performance.

Part II traces the intellectual history of the self-enforcement framework, emphasizing the role of brand names in facilitating contractual performance and contrasting the role of contract terms in the self-enforcement framework with the standard economic view. Part III examines Klein’s contributions to the theory of the firm and vertical integration as a response to the hold up problem. Part IV considers applications of these insights to specific contractual arrangements including: franchising, block booking, vertical restraints, exclusive dealing, and slotting contracts. Part V examines some virtues of the self-enforcement framework as a framework for understanding and analyzing contract law and contract interpretation. Part VI offers some concluding remarks and observations.

## **II. BRAND NAMES, INCOMPLETE CONTRACTS, AND PRIVATE ENFORCEMENT**

The starting point for modern economic analysis of contracts is recognition that the costs associated with contractual specification result in incomplete contracts. The standard economic framework envisioned contracts as written documents that fully defined future performance. To the contrary, most real world contracts are designed with the intent to leave many elements of anticipated performance unspecified but understood.

As is now well recognized in the law and economics literature, contractual incompleteness does not merely result from the mere ink costs of writing down additional terms. Complete contractual specification is prohibitively costly for several reasons unrelated to the costs associated with drafting additional terms. First, complete contractual specification involves wasteful search and negotiation

costs associated with identifying and negotiating contract terms for all potential contingencies. While discovering and negotiating specific responses to unlikely contingencies are likely to have substantial redistributive consequences, and thereby give parties an incentive to expend resources in an attempt to gain advantages over their trading partners during the contract negotiation process, resources devoted to this purely redistributive effort result in a wasteful dissipation of rents. Second, even if the transacting parties were determined to write a fully contingent contract in which a pre-specified response was drafted for each possible future state, the written agreement would still be incomplete because of measurement costs associated with specifying certain elements of performance (for example, the taste of a hamburger) in a legally enforceable manner. Finally, using written terms also imposes the additional cost of rigidity. In other words, because contract terms are necessarily imperfect, transactors can engage in hold-up by rigidly enforcing terms that may be contrary to the parties' intent. This type of hold-up is what occurred in the much discussed Fisher Body-General Motors case discussed in Part III.

The world of incomplete contracts is also the starting point for Klein's analysis of the role of brand names and contract terms in facilitating contractual performance. The fact that many unspecified but understood elements of contractual performance exist, begs the question as to how these elements are enforced. Macauley (1963) and others had demonstrated that reliance on court enforcement of contract terms was rare. A key analytical feature of Klein's theory of contractual performance was that it reconciled both real world phenomena: the prevalence of intentionally incomplete contracts and self-enforcement.

Klein's answer to the question of how these unspecified elements of performance could be enforced was the brand name enforcement mechanism.

Klein and Leffler (1981) derive, and Klein and Murphy (1988) extend, the analysis of the use of privately imposed brand name sanctions to impose a capital loss for the party violating the contractual understanding.<sup>3</sup> The capital loss is created by the future losses associated with the termination of the relationship in the form of the quasi-rents on the nonsalvageable relationship-specific investments that are lost upon termination (the “repeat purchase mechanism”), as well as the loss of reputation in the marketplace caused by the contractual violation (the “reputational mechanism”). The magnitude of the private sanction that can be imposed on a breaching party is defined by the transactor’s brand name capital.

Klein and Leffler, therefore, predict that the amount of brand name capital a party possesses will determine the degree of incompleteness in their contractual arrangements. For example, parties with extremely limited brand name capital will tend to rely on “thicker” contracts where transactors attempt to specify nearly all contingencies in a court enforceable manner. Conversely, “thin” contracts or handshake deals are more likely to be prevalent where substantial brand name capital is present.<sup>4</sup>

The key economic advantage of self-enforcement is flexibility. Brand name-enforced contracts allow transactors to efficiently use information accumulated during the course of the contractual relationship and permits

---

<sup>3</sup> Lott (1988) extends the Klein and Leffler model by introducing random cost or demand shocks. Bull (1987), MacLeod and Malcolmson (1989, 1998), and Levin (2003) present formal models of self-enforcement. For a discussion and other references, see Hermalin, Katz & Craswell (2007).

<sup>4</sup> Brand name capital need not be exogenous. Klein and Leffler (1981) discuss advertising as one method by which firms can invest in increased brand name capital with expenditures on nonsalvageable firm-specific assets. Similarly, transactors might establish and rely upon social networks to enhance the effectiveness of the reputation mechanism. For example, Grief (1993) considers contractual enforcement in the context of social networks that establish a reputation mechanism to prevent opportunistic behavior and facilitate performance in the absence of court enforcement. Hartmann and Gil (2007) demonstrate the role of ethnic and social networks in facilitating contractual performance in the dry cleaning industry.

transactors to more easily modify or breach the agreement. The primary benefit associated with this type of contractual flexibility is that transactors are assured that performance will take place over a broad range of possible ex post market conditions. The cost of this flexibility is the possibility that transactors “hold up” their trading partners by taking advantage of unspecified elements of performance and attempting to appropriate the available quasi-rents resulting from relationship-specific investment.

Both transacting parties consider the potential gain from this type of hold-up from breaching the underlying contractual understanding with the capital loss imposed by the private sanction. The magnitude of the private sanction that can be imposed in the case of non-performance defines what Klein refers to as the “self-enforcing range” of the contractual relationship. While changes in market conditions might increase or decrease the value of specific investments and therefore make hold-up more or less profitable, a hold-up will only occur when the gain from doing so exceeds the private sanction that can be imposed if the breach is detected. When market conditions change in sufficient magnitude to render hold-up profitable despite the privately imposed sanction, the contract is outside the self-enforcing range and hold-up will occur.

Klein’s self-enforcement framework has very different implications for the role of contract terms and contract law than the standard economic approach. The standard economic analysis of contract terms underlying much of the principal-agent and mechanism design literatures envisions contract terms as instruments to create optimal incentives on some court-enforceable proxy for performance.<sup>5</sup> This standard approach to contract theory, sometimes associated

---

<sup>5</sup> Indeed, enforcing contracts as written is the court’s exclusive role in the conventional economic theory of contracts. For a survey of the principal-agent literature, see Hart and Holmstrom (1987).

with the property rights theory of the firm,<sup>6</sup> incorporates contractual incompleteness by allowing for the possibility that rational transactors might omit contingencies because of failures to foresee and provide for those contingencies or the inability of courts to verify performance. This view is often contrasted with what the literature refers to as the “transaction cost economics” (TCE) theory of the firm associated with Klein and Oliver Williamson (1975).<sup>7</sup> Klein describes the shortcomings of standard contract theory as follows:

“The problem with the standard economic framework is that court enforcement and private enforcement are considered as alternatives – firms will rely upon one or the other, but never both. Principal-agent models, for example, formulate the contracting problem as if transactors do not possess any private enforcement capital. Therefore, it is not surprising that these models have limited predictive value in explaining real world contract terms. On the other hand, standard economic models of reputational enforcement provide no role for contractual specification. However, given the fact that private enforcement capital is limited, transactors can be expected to use written contract terms and, hence, the assistance of the court, as a supplement to private enforcement.”

The economics literatures on contracting and self-enforcement have generally proceeded on two parallel tracks without fully considering the complementary relationship between court and self-enforcement and its implications for both contract design and interpretation. The standard approach envisions the use of written terms solely as an instrument to create an incentive to perform with respect to some court-enforced, contractually specified measure of performance. The self-enforcement framework implies a complementarity between court

---

<sup>6</sup> See, e.g., Grossman and Hart (1984), Hart and Moore (1988, 1990).

<sup>7</sup> Gibbons (2005) describes the property rights theory of the firm as the “inverse” of TCE. Where both rely on contractual incompleteness and asset specificity, “[TCE] envisions socially destructive haggling ex post, the property-rights theory assumes efficient bargaining, and where [TCE] is consistent with contractible specific investments ex ante, the property-rights theory requires non-contractible specific investments.”

enforcement and private enforcement, in the sense that the enforcement mechanisms are more effectively used in conjunction with one another than separately. Klein (1996) describes the purpose of contractual specification as an attempt to

[e]conomize on the amount of private enforcement capital to make a contractual relationship self-enforcing by merely “getting close” to desired performance in a wide variety of circumstances (without creating undue rigidity) and to let the threat of private enforcement move performance the remainder of the way to the desired level.

There are two ways contract terms can operate within this framework. The first is to directly control non-performance with written terms. Contract terms can specify court-enforceable measures of performance such as quantity, quality, price, color, or date of delivery. By using written terms to specify performance of these elements in a court-enforceable manner, the parties decrease the amount of private capital necessary to keep the relationship within the self-enforcing range. The second mechanism through which contract terms can broaden the self-enforcing range, and therefore reduce the probability of hold-up, is by shifting private enforcement capital between the parties. In other words, contract terms may be used to ensure that the private enforcement capital aligns with the parties’ relative hold-up potentials, thereby ensuring that the party with the greater hold-up potential has more to lose by non-performance than his trading partner with less hold-up potential.

The use of exclusive territories in distribution relationships is an illustrative example of the major differences in approaches to a common contract term. The standard contract theory view of exclusive territories contemplates the grant of the territory as an attempt to create optimal incentives for franchisee performance as increasing the franchisee’s exposure to repeat customers and therefore internalizing the externalities associated with dealer free-riding on the

franchisor's brand name.<sup>8</sup> Applying the alternative self-enforcement approach, Klein and Murphy (1988) and Klein (1995) demonstrate how the exclusive territory also creates a premium stream for the franchisee and thus increases his costs of nonperformance if the franchisor terminates the relationship. Instead of viewing contract terms as creating the correct performance incentives with respect to some contractually specified measure, the self-enforcement framework implies that transactors will select contract terms to economize on their scarce private enforcement capital and broaden the self-enforcing range to assure performance under the broadest possible range of likely ex post market conditions and thereby increase the parties' ability to flexibly adapt to changes in those conditions. Application of this framework in order to understand how the adoption of a particular contractual arrangement increases the self-enforcing range has particularly valuable in explaining choices about organizational form and contractual choice.

### **III. THE HOLD UP PROBLEM & THE THEORY OF THE FIRM**

Klein's most cited contributions to both economics and law and economics concern the theory of the firm. Kim et al. (2006) list KCA as the 30<sup>th</sup> most cited economics paper from 1970 to 2005 despite the fact that the article was not cited until 1980.<sup>9</sup> The vertical integration of General Motors and Fisher Body, originally offered as a motivating example in support the theoretical framework presented in KCA, has become the paradigmatic example of the theory of the firm in action in textbooks, articles, and classrooms in law schools, business

---

<sup>8</sup> See Rey and Stiglitz (1995), Rey and Tirole (1996) for examples of the standard approach.

<sup>9</sup> This measure likely underestimates the impact of KCA on law and economics for another reason: it does not the impact on legal scholarship and law and economics scholarship in legal journals. A Westlaw search of the JLR database reveals an additional 353 citations.

schools and economics departments.<sup>10</sup> However, Klein's contributions to the theory of the firm are also part of a more general framework for understanding the role of contract terms and self-enforcement. The primary advantage of vertical integration within this framework is the increased flexibility that transactors gain by avoiding the use of a rigid long-term contracts to supplement their reputational capital.

This section focuses on Klein's contributions to the literature examining the boundaries of the firm applying this framework, with emphasis on the relationship between those boundaries and asset specificity. Whinston (2003) has described the theoretical and empirical confirmation of the relationship between asset specificity and vertical integration as "one of the great success stories in industrial organization over the last 25 years."<sup>11</sup> To provide some context for these contributions, I will first discuss the origins and subsequent developments of the literature.

Coase (1937) first identified the relationship between transaction costs and the scope of the firm as a fundamentally important research question for industrial organization economics. Since Coase's seminal contribution, the same question has attracted the attention of many economists and legal scholars attempting to identify the determinants of firm scope. Thus, not surprisingly, there are several non-mutually exclusive "theories of the firm" that emphasize different aspects of the vertical integration decision. The primary theories discussed in the literature are the TCE approach, generally associated with Klein and Williamson, and the modern property rights approach (PRT), developed by

---

<sup>10</sup> A Google search for "Fisher Body" & "General Motors" & "Theory of the Firm" generates 1,310 hits (search conducted 03/06/2008).

<sup>11</sup> The positive relationship between asset specificity and vertical integration is documented in several surveys of the empirical literature in Lafontaine and Slade (2007), Joskow (1988), Shelanski and Klein (1995), Crocker and Masten (1996), Lyons (1996), Coeurdoroy and Quelin (1997), and Masten and Saussier (2000).

Grossman and Hart (1986) and Hart and Moore (1990).<sup>12</sup> While the TCE approach emphasizes the costs associated with ex post bargaining, PRT generally assumes efficient and costless ex post bargaining and emphasizes the choice of ownership structures in providing optimal incentives for ex ante investment.

While the distinction between ex ante incentives and ex post bargaining is somewhat useful in distinguishing the general features of the TCE and PRT approaches, there are also important similarities and more subtle differences that have led to some difficulty in identifying their distinguishing features. These difficulties have led some to re-examine the TCE and PRT approaches to distill their primary features into simple formal models. For example, Gibbons (2005) supplies basic formal models for four leading theories of the firm distilled from the work of Klein, Williamson, Hart, and Holmstrom:<sup>13</sup> (1) a “rent seeking theory,” (2) a “property rights theory” (PRT), (3) an “incentive system” theory, and (4) an “adaptation” theory.

Important similarities exist between the theories focusing on ex ante incentive alignment and ex post contractual governance. For example, the approaches have in common a central role for contractual incompleteness. While the “rent seeking” and “adaptation” approaches emphasize the relationship between that incompleteness and the ability hold up one’s trading partner to increase their share of quasi-rents, PRT emphasizes physical ownership of assets and residual control rights that exist because of contractual incompleteness. Recent literature surveys somewhat incompletely characterize TCE approaches as emphasizing ex post contracting costs and largely ignoring ex ante incentives,

---

<sup>12</sup> See also Hart (1995), Holmstrom & Milgrom (1991, 1994), Holmstrom & Tirole (1991) and Holmstrom (1999) for some key contributions to the PRT approach.

<sup>13</sup> See also Whinston (2003).

while PRT approaches focus exclusively on ex ante incentives to invest and assume costless renegotiation.<sup>14</sup>

The “rent-seeking” and “adaptation” theories of the firm were developed by KCA (1978), Klein (1996, 2000, 2007), Klein and Murphy (1988, 1997), and Williamson (1971, 1985), and emphasize the role of integration in preventing socially destructive “haggling” over “appropriable quasi-rents.” While these simplifying labels come with some risk of obfuscating important and subtle differences between theories, they are useful for highlighting some of the critical features of the theories. For example, the “rent seeking” label correctly captures the fact that a hold up involves attempts to redistribute wealth between parties, and that the resources parties expend in attempts to obtain and prevent these transfers also have allocative effects. Of course, the view that the TCE approach focuses exclusively on ex post contracting costs is overstated, as Klein (1996) and others have also emphasized ex ante contracting costs, where transactors engage in a wasteful search for an informational advantages over transacting partners during the negotiation process.<sup>15</sup>

Klein (1996, 2000, 2007) and Klein and Murphy (1988, 1997) also emphasize the role of vertical integration in facilitating coordination by reducing contractual specification, thus reducing reliance on court enforcement to enforce intentionally incomplete, imperfect, and rigid long-term contracts.<sup>16</sup> Gibbons (2005) calls this relationship between vertical integration and ex post contractual

---

<sup>14</sup> Lafontaine and Slade (2007) provide a very useful survey of the empirical evidence with respect to these theories. Gibbons (2005) focuses on theoretical contributions.

<sup>15</sup> See Schmitz (2007) for a recent attempt to include inefficient rent-seeking into a property rights model.

<sup>16</sup> Williamson (1971, 1973, 1975, 1991) also emphasizes the role of vertical integration in improving sequential decision-making as uncertainty is resolved over time. While Williamson (1975) focuses on the relationship between integration and relational adaptation within the firm, Gibbons (2005) distinguishes Klein’s work as emphasizing the role of relational contracting between firms.

adjustment the “adaptation” theory of the firm. The coordination advantages of integration within this framework stem largely from the avoidance of “rigidity costs,” including potential hold-up problems associated with rigid and imperfect long-term contracts. The costs saved by using integration rather than contract to avoid hold up problems are not limited to negotiation and “ink” costs. Rather, vertical integration avoids costs associated with using explicit terms in a long-term contract when market conditions are uncertain over time. This advantage in coordinating economic activities through integration, in turn, depends on the conditions that must be present for using long-term contracts in the first instance. For example, specific investments and insufficient reputational capital must be present because the parties could otherwise coordinate activities without long-term contracts.

The most well known example of the rigidity costs associated with contractual specification is the Fisher Body-General Motors case first discussed briefly in KCA (see also Klein 1988, 2000). The Fisher Body-General Motors case has become a classic example in economics illustrating the relationship between asset specificity and vertical integration, a relationship that has been demonstrated empirically in many settings and industries. However, the primary economic significance is that Fisher Body and General Motors operated with a long-term exclusive dealing contract before that contractual arrangement was replaced in favor of vertical integration, and thus highlights the relative advantages and disadvantages of each arrangement.

There has been a great deal of subsequent discussion over the details of the relationship between Fisher Body and General Motors and the events that led to integration. Coase (2000, 2006), Freeland (2000), and Casadesus-Masanell and Spulber (2000) have offered alternative accounts of the integration story and argued Fisher Body did not hold up General Motors, integration was motivated

by other considerations, and long-term contracts are always sufficient to resolve hold-up concerns. In some ways, it is difficult to know what to make of this literature challenging Klein's (1978, 1988, 2000) account of these events. Some of these accounts appear to challenge the entire empirical literature demonstrating the relationship between asset specificity and vertical integration by attacking this single example designed to illuminate how the mechanism could work in practice by generating greater market contracting costs, thus leading to greater vertical integration. After 30 years of debate about this particular example, this much is clear: Klein's theory of hold up, and its primary prediction of a positive relationship between asset specificity and vertical integration had held up quite well and has become one of the most well-documented economic phenomena in industrial organization economics.<sup>17</sup>

But this critical literature, and Klein's responses to it, has also served a productive purpose. The details of the Fisher Body hold-up are critical precisely because they provide an opportunity to understand the mechanism behind the failure of the long-term exclusive dealing contract in favor of vertical integration. This critical literature led by Coase has produced a better understanding of the events that led to Fisher Body's hold up of General Motors and its execution. Indeed, the set of critical papers from Coase (2000, 2006) and others have led Klein (2000, 2007) to extend the analysis of the Fisher Body-General Motors case. For example, Klein (2007) provides additional evidence from the actual 1919 Fisher Body – General Motors contract, previously unavailable, to add to the substantial evidence of the economic relationship compiled in the Du Pont case record and other sources.<sup>18</sup> After 30 years of study and debate, the primary

---

<sup>17</sup> See empirical literature cited *supra* n. 11.

<sup>18</sup> Klein (2007) demonstrates that Fisher Body did not hold up General Motors by actually mislocating plants or adopting an inefficient low-capital intensive production technology. While

lesson of the Fisher Body-General Motors remains the same: transacting parties choose contract terms to economize on the reputational capital necessary to make the contractual relationship self-enforcing under the widest possible range of ex post circumstances. But, as the Fisher Body-General Motors' relationship illustrates, the contract terms parties choose can also result in additional hold-up potential.

Klein's emphasis on the role of contract terms within the self-enforcement framework is an important distinction between the contributions of Klein and Williamson.<sup>19</sup> While both adopt the TCE approach and have made fundamentally important contributions to both the theory of the firm and contractual choice, Klein's self-enforcement framework has generated a lens through which specific contractual arrangements can be examined to increase our understanding of how those terms broaden the self-enforcing range. The primary advantage of vertical integration within this framework is the increased flexibility transactors gain by avoiding the rigidity associated with long-term contracts. The framework is useful to explain vertical integration as a particular contractual arrangement that facilitates self-enforcement rather than a special case, and can be applied to explain a number of other contractual arrangements. Indeed, Klein (1996) notes that the self-enforcement framework "should be

---

Fisher Body did threaten to mis-locate plants, Klein documents how Fisher Body accomplished its hold up of General Motors by "negotiating a highly favorable contractual adjustment in 1922 in response to General Motors' demand that Fisher Body's substantial new investment in additional capacity take the form of co-located body plants." This contractual adjustment led to the previously observed substantial decrease in Fisher Body's capital-to-sales ratio. This explanation reconciles the existing evidence concerning threats of plant mis-location, the reduction in capital-to-sales ratio, and the contractual adjustment. Coase (2006) does not recognize that the 1922 contract renegotiation was made to overcome Fisher Body's threatened refusal to co-locate its plants.

<sup>19</sup> A second difference involves some differences in the definition of hold-up. Williamson's (1985) definition of a hold-up includes moral hazard behavior, dishonesty and surprise, whereas Klein (1996) argues that moral hazard is a form of contractual non-performance, not hold-up, because it is fully anticipated.

judged by how much it assists us in explaining the particular contractual arrangements in the marketplace.” The fruits of Klein’s application of the self-enforcement framework are discussed in Part IV.

#### **IV. THE SELF-ENFORCEMENT FRAMEWORK AND CONTRACTUAL CHOICE**

The self-enforcement framework entails the combination of insights discussed in Parts II and III – the threat of loss of future sales and reputation and the role of contract terms in controlling hold-up behavior – and provides a productive framework for analyzing contractual arrangements. This section’s primary purpose is to survey the application of this framework to a number of contractual arrangements.<sup>20</sup>

It is worth repeating that the role of contract terms within this framework is significantly different from the standard economic approach in ways critically important to the economic analysis of contract law.<sup>21</sup> The fundamental motivation for using court-enforced contract terms in the self-enforcing framework is to supplement self-enforcement. If sufficient reputational capital was available, transactors would generally choose to avoid the costs of contractual specification by relying exclusively on self-enforcement. However, because such capital is generally limited, parties must supplement the self-enforcement mechanism with written terms.

The self-enforcement framework generates two fundamental insights about contracting. The first is about the very nature of contract terms. Within the self-enforcement framework, contract terms may be thought of as accomplishing one of two goals: decreasing short-term gains achieved by not

---

<sup>20</sup> See also Masten (2000).

<sup>21</sup> See *infra* Part V for a discussion of the advantages of the self-enforcement framework view of contract terms for contract interpretation.

performing in a manner consistent with the contractual understanding (W1), or increasing the transactor's reputational capital (W2), the capital cost of the lost expected future profit stream imposed upon a nonperforming transactor upon termination. While the economic literature generally recognizes the role of contract terms in defining some element of performance, and therefore reducing W1 by controlling this element through court-enforcement, within the self-enforcement framework the role of contractual specification is to render the residual W1, after all cost-effective contract terms are specified, less than W2. What is less well-recognized in both the economics and law and economics literatures is that contract terms may be used to shift expected future rents, and therefore reputational capital, between transactors. The economic rationale for utilizing contract terms to shift future rents is to more closely align each transactor's reputational capital with his potential expected gain from non-performance. Some examples of the use of contract terms in this spirit are explored throughout this section.<sup>22</sup>

The second fundamental insight, explored in greater detail in Part V, is that the self-enforcement framework demonstrates a fundamental complementarity between court-enforcement and self-enforcement. Klein (1996) describes this complementarity as follows:

The two enforcement mechanisms are substitutes in demand, in the sense of a positive cross elasticity of demand, so that an increase in the price of one mechanism leads to an increased use of the other mechanism. But the two mechanisms are complements in supply, in the sense of a positive cross elasticity of supply, so that an increase, for example, in the quantity of reputational capital leads to an increase in the marginal productivity of court enforcement. That is, the two mechanisms work better together than either of them do separately.

---

<sup>22</sup> See Goldberg (1976), Goldberg & Ericson (1987) and Crocker & Masten (1991) on the use of contract terms to facilitate self-enforcement and contractual flexibility including the adjustment of prices as market conditions change over time.

This complementarity has important implications for the role of contract law and contract interpretation in broadening the self-enforcing range of contracts.

The self-enforcement approach can be usefully contrasted with the contract design literature that I have described as the “standard” or “conventional” approach. This description is not meant to be entirely critical. This literature has been highly influential in economics, heavily cited by law and economics scholars, and supplied some important insights about contracting behavior.<sup>23</sup> However, it cannot be avoided that this conventional approach to contracts envisions a different role of contract terms in assuring performance than the self-enforcement approach. Specifically, this literature examines the role of specific contract terms in minimizing malincentives, given that performance can only be contracted upon imperfectly. While this literature takes incomplete contracting very seriously, and has produced several fundamentally useful insights about contracting, it has the drawback of generally envisioning agreements as either court-enforced or self-enforced, but not both. In other words, the standard contract theory literature contemplates the purpose of contracts as fundamentally about creating optimal incentives for performance on some imperfect, but court-enforceable, measure.

The similarities and differences between these approaches should be apparent. Both approaches are anchored by a foundation which accepts the notion that contracts are incomplete. Both approaches also acknowledge the role of contracts terms in reducing the gains from non-performance that derive from contractual specification. But the mechanism through which specification affects performance is quite different between the two literatures. In the standard literature, specification can prevent high-value opportunism by using contract

---

<sup>23</sup> See, e.g., Posner (2003); Smith and King (2007).

terms to provide direct incentive to perform where non-legal sanctions are insufficient to facilitate performance. Alternatively, the self-enforcement approach models the contracting process as an attempt to use court-enforced terms as a *complement* to limited reputational capital in order to render an economic relationship self-enforcing over the broadest possible range of ex post market conditions. The standard approach leaves no room for the interaction between specification and reputational capital. Competing economic models should be judged by their predictive power of real world contractual arrangements. By this measure, the self-enforcement mechanism has been highly successful, whereas the principal-agent literature has been underwhelming in its ability to explain the content of contracts.<sup>24</sup>

The Fisher Body - General Motors exclusive dealing contract again illustrates the predictive power of the self-enforcement approach. Fisher Body had entered a long-term, cost-plus exclusive dealing contract. The purpose of this contract was to encourage Fisher to make General Motors-specific investments and protect Fisher from hold-up by locking General Motors into Fisher. Instead, when market conditions changed dramatically, Fisher Body was able to hold-up General Motors and negotiate a significant side payment, and the parties eventually vertically integrated. The standard view cannot explain the original exclusive dealing/ cost-plus arrangement the parties adopted, as they created a direct incentive for Fisher to increase the cost of auto bodies. However, the self-enforcement approach demonstrates why the terms Fisher Body and

---

<sup>24</sup> See, e.g., Posner (2003) ("so far the literature has failed to predict the content of either contracts or legal doctrines"). Posner points out that this literature, which he overbroadly refers to as the literature on incomplete contracts without referencing work on self-enforcement, predicts contracts that are much more complex than designed by real parties. Posner explains this gap between theory and evidence by suggesting that real-world parties are less rational than those in the economic literature. However, the real world observation of contracts less complex than those predicted by the principal-agent literature is also fully consistent with the self-enforcement approach, where contract terms are used to supplement reputational capital.

General Motors adopted were efficient when adopted and operated successfully for several years without Fisher Body attempting to hold-up General Motors.

In addition to vertical integration, Klein has applied the self-enforcement framework to explain the role of a number of contractual arrangements with significant success over the past 30 years. These contributions have been highly influential in the literatures, as well as with judges. For organizational purposes, these contributions will be divided into two broad categories: (1) vertical restraints; and (2) franchising.

### **A. THE ECONOMICS OF VERTICAL RESTRAINTS**

Perhaps the most influential of these contributions has been the explanation of vertical restraints between manufacturers and retailers. For example, Klein and Murphy (1988) and Klein (1999) demonstrate the role of resale price maintenance (RPM) and exclusive territories in facilitating contractual performance and supply of the desired dealer services. Manufacturers use vertical restraints to decrease the short-run gain from cheating ( $W1$ ) by limiting the ability to expand output and increase the long-run gain from performance ( $W2$ ) by creating a quasi-rent stream that will be lost if non-performance is detected.<sup>25</sup> This explanation of RPM in providing a premium stream to dealers in order to induce the supply of additional promotional services has successfully provided a pro-competitive explanation of RPM where the standard explanation of inter-dealer free-riding does not apply (Telser 1960). Indeed, Klein and Murphy's (1988) pro-competitive explanation of RPM as a method to facilitate the supply of promotional services was recently relied upon

---

<sup>25</sup> Klein and Murphy (1988) demonstrate how Coors implemented a combination of RPM with an exclusive territory to create a premium stream and encourage dealers to provide refrigeration and product rotation services for Coors beer that underwent a brewing process that did not include pasteurization and deteriorated quickly at room temperature.

by the Supreme Court in *Leegin Creative Leather Product, Inc. v. PSKS, Inc.*, in overturning the longstanding per se prohibition against minimum RPM.<sup>26</sup>

Klein and Murphy's analysis of vertical contractual relationships also provides another fundamental insight concerning a pervasive incentive conflict between manufacturers and retailers with respect to the supply of promotional services. This insight has been the foundation for a number of analyses of contractual arrangements in the retail setting. Specifically, Klein and Murphy demonstrate that retailers will undersupply promotional services because manufacturers do not take into account the incremental profit margin earned by the manufacturer on promotional sales when some, but not all, consumers value the promotional service. RPM or other vertical restraints can create a premium stream for the retailer to induce provision of promotional services that would not otherwise be supplied under the threat of termination for non-performance. Retailers that fail to provide those services are "free-riding" on the contractual arrangement by taking the compensation and failing to deliver the contracted for services.<sup>27</sup>

Consider the case of promotional shelf space. Klein and Wright (2007) apply the insight of the Klein and Murphy model to explain slotting contracts, shelf space contracts between manufacturers and retailers. Klein and Wright illustrate that when deciding how much promotional shelf space to provide a manufacturer's product, retailers will not take into account the manufacturer's

---

<sup>26</sup> In addition to the Supreme Court's reliance on the Klein and Murphy analysis of vertical restraints and RPM in *Leegin*, the analysis has also been widely cited in the legal literature for its implications for antitrust law, franchising, and contract law, including 77 citations in the JLR database.

<sup>27</sup> This type of free-riding should be distinguished from what Klein refers to as the "classic retailer free-riding" analysis popularized by Telser (1960). As Klein (1988, 1993) and others, such as Potosfky (1983), have pointed out, the services full-service retailers provide in cases involving RPM generally do not involve this type of free-riding. Rather, the promotional services provided are those for which the free-riding consumers are unwilling to pay.

profit margin on the incremental sales produced by the promotional shelf space. This problem is particularly significant when the manufacturer supplies a differentiated product where the wholesale price it receives is substantially greater than its marginal production cost. Thus, retailers will not find it in their interests to supply the promotional shelf space necessary to generate the manufacturer's profitable incremental sales. Manufacturers, therefore, must find a way to contractually incentivize retailers to supply their products the desired promotional shelf space. In the grocery retail setting, Klein and Wright show that the efficient form of compensation often involves a lump-sum per unit time slotting fee when inter-retailer competition on the particular product makes compensation with a lower wholesale price a more costly way to generate equilibrium retailer shelf space rents.

More generally, Klein has extended the analysis of the incentive conflict between manufacturers and retailers concerning promotional effort in a series of papers examining the role of exclusive dealing terms in distribution contracts. One theme that emerges from these recent analyses is that our current understanding of the potentially procompetitive role of exclusive dealing in preventing dealer free-riding is unduly narrow. For example, Klein & Lerner (2007) offer an expanded economic analysis of free-riding that begins from the premise, discussed above, that manufacturers often want their dealers to supply more promotional effort than the dealers would independently provide. The contractual solution to this incentive conflict is to adopt a marketing arrangement under which the manufacturer compensates the retailer for these extra services. This can take place through the use of exclusive territories, RPM, slotting allowances, or other promotional payments. However, the creation of the necessary premium stream generated by this contractual solution can cause other problems associated with retailer free-riding on the marketing

arrangement. Klein and Lerner (2007) identify three types of free-riding strategies to supplement the conventional economic analysis of retailer free-riding. Specifically, the retailer can “free-ride” on the investments the manufacturer provides to sell rival products, by using the promotional efforts the manufacturer paid for to sell rival products, or by simply not supplying the increased level of brand-specific promotion for which the manufacturer paid.

Free-riding on manufacturer supplied investments to sell rival products is the most frequently discussed in antitrust law and economics and by the courts. Marvel (1982) first identified the rationale for exclusive dealing contracts in preventing this sort of free-riding, which has been accepted in antitrust jurisprudence on the grounds that it promotes valuable investments.

Klein and Lerner (2007) expand on this analysis by debunking economists’ claim that exclusive dealing could not prevent the other two types of free-riding. Indeed, exclusive dealing contracts can also prevent free-riding on paid for promotional efforts, even absent manufacturer supplied-investments that dealers can use to sell alternative products. When a manufacturer solves the promotional incentive conflict by compensating the dealer for providing additional promotional effort with an extra profit margin on the dealer’s sales of the manufacturer’s products, the dealer has the incentive to free-ride by switching their promotional efforts to alternative brands.

Further, Klein and Lerner (2007) present an economic rationale for the commonly claimed role of exclusive dealing in “increasing dealer loyalty” by producing dedicated dealers, even absent any dealer switching behavior. While courts have long accepted the conventional wisdom that exclusive dealing could increase dealer loyalty, little was known about how exclusivity could induce greater promotional effort. Klein and Lerner (2007) demonstrate that exclusive dealing can prevent a third type of dealer free-riding, failing to supply paid for

promotional effort, by increasing the marginal incentive for dealers to promote the manufacturer's products. Klein and Lerner also demonstrate that the exclusive dealing contracts in *Dentsply* were, at least, partially explained by a desire to prevent this third type of free-riding, and that the Third Circuit mistakenly relied upon the conventional economic wisdom in rejecting this rationale.

Separately, Klein and Murphy (2008) have shown that exclusive dealing can "intensify" competition for distribution when a retailer commits to exclusively or primarily promote a single manufacturer's products. The fundamental economic insight of Klein and Murphy's analysis is that the exclusive dealing contract increases the elasticity of demand facing manufacturers bidding for the exclusive shelf space and access to the retailer's customer base. Thus, the competitive bidding process results in increased dealer compensation that, in turn, benefits consumers when payments are ultimately passed on to final consumers.<sup>28</sup> While exclusive dealing contracts inherently make some consumers worse off by reducing product variety within the store, Klein and Murphy demonstrate that competitive retailers enter these exclusive arrangements only when it allows them to compete in the marketplace by offering a combination of price and product variety which will be preferred by a sufficiently large group of consumers. In other words, failing to offer consumers a combination of price and variety in a competitive retail marketplace is unlikely to be profitable under these conditions because retailers will lose a substantial number of consumers and sales to rival supermarkets.

Klein and Murphy (2008) also consider a variety of contractual alternatives designed to intensify manufacturer competition for shelf space while

---

<sup>28</sup> Wright (2001) provides evidence that competition in the grocery retail sector has remained intense despite substantial increases in concentration over the past several decades. Klein and Wright (2007) provide evidence that retailer profitability has remained constant over this same time period during which slotting contracts and payments to supermarkets increased dramatically.

accommodating consumer demand for product variety. These alternatives include “partial exclusives” such as preferential shelf space contracts, market share contracts, and category management contracts, which delegate shelf space allocation input to a manufacturer. A reasonable economic interpretation of the implicit understanding under which a category captain operates is that it will stock other highly demanded brands, in addition to its own products, so as not to unnecessarily reduce overall retailer demand and profitability.<sup>29</sup> Collectively, the insights from Klein’s recent work on the economics of exclusive dealing demonstrate that exclusive dealing, to use antitrust terminology, frequently involves “competition on the merits” and is part of the normal competitive process.<sup>30</sup>

Klein’s analysis of vertical restraints extends beyond contractual provisions facilitating retailer supply of promotional effort. For example, Klein and Kenney’s (1983) seminal analysis of the block booking arrangement prevalent in the American motion picture industry demonstrates that block booking terms were used to reduce ex post opportunism. Klein and Kenney (1983) challenged the prevailing, and still influential, view that block booking arrangements were a form of price discrimination. Stigler (1963) had argued that block booking was a method distributors used to capture additional consumer surplus by price discriminating to exploit the variance in individual film valuations across cities. However, Klein and Kenney challenged this view by presenting evidence from the *Loew’s* case suggesting that block prices were not

---

<sup>29</sup> Wright (2009) applies the insights of Klein and Murphy’s analysis of partial exclusives to the adoption of category management contracts in the retail sector. Specifically, Wright analyzed the competitive consequences of United States Tobacco’s category management arrangement and implementation that were the basis for the *Conwood Co. v. United States Tobacco Co.* antitrust litigation.

<sup>30</sup> See Klein (2003) and Klein (2001) discuss the use of exclusive dealing contracts in the *Microsoft* litigation.

uniform and varied across geographic markets. In addition, using a competitive auction mechanism to grant a single station in multiple station markets the exclusive rights to film distribution is inconsistent with the price discrimination motive, since the auction would reveal individual station values without resorting to block booking.<sup>31</sup>

Kenney and Klein then applied the self-enforcement framework to understand the use of block booking contract terms in the motion picture industry. Specifically, Kenney and Klein argue that block booking prevents buyers from rejecting part of a package of products that is average-priced. The economic rationale for average-pricing in the motion picture industry is to deter opportunism in the form of investments in informational advantages with little or no social value. For example, in *Paramount*, block booking was used to prevent buyers from picking through a subset of films after initial exhibition results became available.<sup>32</sup> Similarly, the block booking arrangement in *Loew's* discouraged wasteful searching for exceptional film values.

Klein and Kenney also apply this framework to De Beers' marketing arrangement of uncut diamonds. De Beers pre-sorted diamonds into relatively homogenous categories to be sold in pre-selected blocks to pre-selected buyers. These pre-selected blocks were to be sold at non-negotiable prices and buyers' rejection of the sales offer would result in the withdrawal of future invitations to the buyer to purchase stones. In the De Beers arrangement, block booking prevents wasteful oversearching to identify the highest quality stones and leaving only low quality stones behind. These pre-purchase investments lead to

---

<sup>31</sup> Kobayashi (2005) discusses the Klein and Kenney analysis as it relates to the economic literature on bundling.

<sup>32</sup> Klein and Kenney (2000) provide evidence of opportunistic exhibitor rejection of films where exhibitors would cancel or shorten the run of "overpriced" films and substitute films from competing distributors.

wealth transfers between consumers and sellers but have little social value. Block booking prevents the wasteful costs of individually pricing the stones in each quality group and is enforced by a brand-name enforcement mechanism. Specifically, De Beers discourages buyers from “cheating” on the arrangement by rejecting sights of lower than average quality by providing buyers a premium stream that exceeds the rents from such rejection. The enforcement mechanism thus relies on De Beers’ brand name because buyers expect they will receive a diamond sight from the underlying quality distribution upon which the average price is based. Similarly, because buyers will lose the premium stream associated with De Beers’ pricing if they reject a sight that is less than average quality, the arrangement is self-enforcing.

Two overarching themes emerge from Klein’s work on vertical restraints. The first is a commitment to increasing our economic understanding of these arrangements as they operate in the real world rather than assuming away the important institutional details of the arrangements for the purposes of analysis. For example, Klein’s analysis of vertical restraints builds on the fundamental economic reality that the normal competitive process often results in manufacturers compensating retailers for the provision of shelf space and other promotional services. This insight builds upon what is a departure from the standard view of the economics of retail in the literature, where retailers merely passively transmit exogenous consumer preferences, and its only economic function is to reduce search costs. To the contrary, Klein’s analytical framework begins in a world where the retailer operates in a competitive environment facing a downward sloping demand, and has some discretion over its shelf space, promotion, and pricing decisions. Without building these two empirical regularities into the analytical framework, it would be nearly impossible to understand the fundamental incentive conflict over the provision of retail

promotion such as shelf space, and therefore also impossible to understand the use of exclusive and partial exclusive contracts in this setting. The adoption of a more realistic framework for analysis that includes brand names, self-enforcement, and a more realistic account of retail competition allows for a more robust understanding of the way these contract terms operate and their competitive consequences.

## **B. THE ECONOMICS OF FRANCHISING**

Franchising is one of the most studied forms of organization in the economics literature. At least legally, franchising is a unique form of contractual organization. For example, franchising is the subject of a thicket of state and federal regulations, whereas analytically similar contractual arrangements are not subjected to special scrutiny. On the other hand, franchising is just one of many distribution arrangements where manufacturers exercise control over their exclusive retailers where these economics forces are present. Franchising has provided a set of arrangements for understanding the role of contract terms in reducing transactions costs and incentive conflicts.

The underlying economics of franchising are similar to the economics of vertical restraints. The interesting economic questions concerning franchising are similar to those concerning contractual choice in vertical relationships: what is the role of the particular contractual terms adopted in the franchise contract, and what determines whether a franchise is owned and operated by the franchisor rather than owned and operated by an independent franchisee?

I will focus on three of Klein's key contributions to the law and economics of franchising: (1) application of the promotional incentive conflict analysis originally articulated in Klein and Murphy to franchising arrangements; (2) explaining contract terms franchisors and franchisees adopt to facilitate self-enforcement; and (3) applying the economic analysis of franchising to antitrust

aftermarket claims to establish the legal and economic principle that market-power should be measured ex ante at the time of contracting rather than ex post.

The first economic question with respect to franchising is to understand why franchisor and franchisee incentives do not coincide and thus require contractual restraints. If the incentives for performance were aligned, there would be no reason for contractually controlling franchisee behavior. A number of reasons have been studied at great length in the economics literature. For example, when franchisees jointly use a common brand name, each franchisee can free-ride on the brand name by reducing quality, because the quality reduction results in an “across-the-board” reduction in future demand.<sup>33</sup> A second, well-known form of free-riding, discussed above, involves franchisee supply of some pre-purchase service that consumers can get free of charge at a full service franchisee before purchasing at a presumably lower price from a “free-riding” franchisee who does not provide the services. This is what Klein has referred to as the “special services” free-riding problem originally articulated by Telser (1960). A third form of incentive conflict involves “successive monopoly” or “double marginalization.” Where the franchisee possesses some power over price, it is sometimes claimed that franchise restraints are an attempt to control the franchisee’s incentive to increase price to reflect its market power.

A critical and distinguishing feature of Klein’s analysis of franchising agreements is its focus on a fourth incentive conflict: the supply of promotional services. This incentive conflict is much more general and pervasive than the three “special case” scenarios tend to suggest. All that is required, as discussed above in the context of vertical restraints, is that the franchisee control provision

---

<sup>33</sup> See Rubin (1978) and Klein (1980). Rubin (1978) provides an early transaction-cost economics-based explanation of franchising based upon the costs of monitoring employee-operated franchise outlets.

of some inputs that influence demand for the franchisor's product, and that the franchisor sells those products at a mark-up over its marginal cost. Attentive readers will recognize these conditions as analogous to those articulated by Klein and Murphy (1988) and Klein and Wright (2007) with respect to vertical restraints. The emphasis of this fourth promotional incentive conflict appears to explain the prevalence of vertical restraints in settings where special services free-riding, the use of common brand names, or successive monopoly problems are absent or insignificant.

Using the promotional malincentive problem as a starting point for the analysis, Klein applies the self-enforcement framework, discussed in greater detail above, to analyze the role of contract terms in franchise contracts. Again, the critical economic point is that contract terms can operate in two ways to facilitate performance. The first method is to provide the correct marginal incentive for franchisee performance by contractually specifying the desired performance to be enforced by a court. Contractual specification alone is generally insufficient to assure performance because it is impossible, or at least prohibitively costly, to fully specify a legally enforceable agreement documenting each element of performance. The second method contract terms can facilitate performance is by creating sufficient franchisee rents so the threat of termination by the franchisor generates sufficient incentive to supply the desired behavior. The self-enforcement mechanism, as discussed in Klein and Leffler (1981) and Klein and Murphy (1988), requires the franchisor to monitor franchisee behavior and terminate the franchisee for nonperformance. The important distinction is that the threat of termination, and not the threat of litigation to enforce the contract terms, produces the incentive to perform.

As with the Klein and Murphy (1988) analytical framework for vertical restraints, this view of franchising contract terms raises the question of how these

terms create direct performance incentives or economic rents. For example, Klein and Murphy (1988) and Klein (1995) discuss the use of exclusive territories, RPM, and the number and spacing of outlets as alternative ways to create the franchisee premium stream.<sup>34</sup>

A second important economic question related to franchising involves when franchisors will rely on company ownership as opposed to the franchising of outlets. Economists have made significant efforts to theoretically and empirically identify the determinants of company owned outlets and the content of franchise contract terms.<sup>35</sup> The self-enforcement framework also implies that the key determinant of vertical integration of outlets is the monitoring costs of franchisor-owned outlets relative to franchised outlets. This prediction is consistent with the empirical finding that state franchise laws prohibiting franchisee termination except for “good cause” result in reduced welfare. For example, Klick, Kobayashi, and Ribstein (2007) find a positive relationship between state franchise termination restrictions and the fraction of franchisor-owned outlets, and a negative relationship between the regulations and the overall level of franchising.<sup>36</sup> Brickley and Dark (1987) also identify a negative relationship between distance of an outlet to the nearest franchise headquarters and franchisor ownership, though it is unclear whether this evidence strongly supports the relationship between franchisee-monitoring costs relative to the costs of monitoring a franchisor-owned outlet.<sup>37</sup>

---

<sup>34</sup> Kaufmann and Lafontaine's (1994) provide evidence that McDonald's franchisees retain significant rents.

<sup>35</sup> See, e.g., Brickley & Dark (1987); Brickley, Dark, & Weisbach (1991b); Bhattacharyya & Lafontaine (1995); Lafontaine (1992); Lafontaine & Shaw (1999).

<sup>36</sup> See also Brickley et al. (1991); Beales & Muris (1995); and Smith (1982).

<sup>37</sup> Klein (1995) argues that distance of an outlet to the nearest franchisor headquarters is not a reasonable proxy for monitoring costs because if they are significant at all, these costs would also increase the costs of monitoring franchisor-owned outlets.

A third key feature of Klein's analysis of franchise arrangements is its application to antitrust analysis of franchise tying arrangements. In the typical franchise tying claim, a franchisee challenges an underlying requirements contract as an illegal tie-in under the Sherman Act. A necessary element of a franchise tying claim was that the franchisor possessed antitrust market power. Klein and Saft (1985) first recognized the key economic insight involving antitrust analysis of these franchise tying cases: that franchisor market power should be measured, contrary to the legal presumption after *Siegel v. Chicken Delight*, during the pre-contract time period when franchisees make their decisions and competition between franchisors is reflected in contract terms. Klein (1999) motivates this analysis with the example of a law firm looking for office space to rent. From the perspective of the law firm contemplating a lease, and before making significant investments into its new office space, there is significant competition among potential building owners, as many close substitutes exist for the space the law firm is seeking. Building owners cannot and do not exercise power because of the competition between owners at the pre-contract stage. Therefore, Klein concludes, the presence of pre-contract competition can be expected to generate a competitive package price for the space and other contract terms such as parking rates, air conditioning rates, and other amenities.

The "hold up" problem at the root of most franchise tying cases occurs only after the law firm has signed a long-term lease, made investments, and becomes "locked-in." Prior to the Klein and Saft (1985) analysis, the legal presumption was that this type of post-contract hold-up was a monopoly problem associated with market power. But consider again the law firm who entered into a competitively negotiated package price. Defining market power ex post in such a case would necessarily mean that market power existed in each

scenario, even when contract terms were negotiated in highly competitive markets where buyers make sunk investments. Similarly, Klein and Saft (1985) and Klein (1999) persuasively argued that antitrust laws associated with monopoly power should measure market power at the pre-contract stage. In sum, although the franchisor's ability to "hold-up" the franchisee might result in bad outcomes for the franchisee, not all bad outcomes in all contracts are antitrust problems. Antitrust problems require the exercise of monopoly power, whereas the typical franchise tying case does not. Conflating contract problems associated with imperfect buyer information or hold-up potential with the exercise of monopoly power threatens to cast the shadow of antitrust liability and treble damages over mundane contractual disagreements.

This insight has been extremely influential in antitrust analysis of franchise tying arrangements and aftermarket tie-in claims. For example, at least 9 federal courts have cited the Klein and Saft analysis in rejecting franchise tying claims, including the Supreme Court in its landmark 1992 *Kodak* decision. While the *Kodak* decision was widely thought to have opened the door to "aftermarket" franchise tying claims, Goldfine and Vorrasi (2004) demonstrate, lower federal courts have all but completely rejected the invitation from *Kodak* to expand "aftermarket lock-in" theory in the franchise setting or elsewhere. Courts commonly reject these theories on based on the distinction, borne in Klein's analysis, between "contract power" and "antitrust power." This sensible development in antitrust treatment of tying arrangements is owed in large part to the strength of the Klein and Saft (1985) and Klein (1999) arguments that (1) hold-up is a problem distinct from the exercise of monopoly power, (2) antitrust market power must be present at the pre-contract stage, and (3) transactors aware that they are entering an agreement without complete information and with the potential for hold-up will make arrangements to protect themselves

(either with contractual terms or by dealing with firms with large reputations and much to lose by executing a hold-up). It should be pointed out that despite popular descriptions of *Kodak* as endorsing post-contract evaluation of antitrust market power, Klein (1999) correctly points out that *Kodak* approved the aftermarket hold-up logic only under limited circumstances include the existence of an actual hold-up. The limitations adopted by the *Kodak* court carefully avoid the expansion of antitrust market power to all contracts signed in competitive markets where buyers have made sunk investments and therefore some modicum of hold-up potential exists. These limitations along with the death of franchise tying and *Kodak* aftermarket claims in the lower courts suggests that Klein's work in emphasizing the distinction between the economic forces associated with hold-up and those associated with antitrust market power have been highly influential in altering the path of antitrust jurisprudence in this area.

An additional key contribution of Klein and Saft (1985) is the economic explanation of tie-ins in economizing on monitoring costs of franchisees. For example, Klein and Saft demonstrate that a franchise tie-in might reduce the common brand name free-riding because it reduces the costs of policing and detecting franchisee cheating since the franchisor can identify cheating whenever he sees a non-authorized product in the franchisee outlet, rather than making more difficult quality comparisons. Klein and Saft also show that franchise tie-in arrangements might efficiently meter value across franchisees by placing an upcharge on inputs, rather than collecting the franchise fee on sales.<sup>38</sup> Klein and Saft (1985) conclude their analysis with the following observation concerning the potential implications for antitrust law: "an understanding of these pro-

---

<sup>38</sup> Klein (1999) argues that the requirements contracts in *Chicken Delight* and *Kodak* are likely efficient components of a discriminatory metering arrangement. Klein (1999), Klein (1995) and Klein and Wiley (2001) argue that contractual arrangements that foster economic price discrimination enhance efficiency and should be immunized from antitrust liability.

competitive benefits is necessary before the court is likely to adopt an explicit rule of reason standard, rather than the extremely rough de facto rule of reason standard we now have.” While the Court has not taken the opportunity to overturn *Jefferson Parish* and explicitly adopt the rule of reason approach, lower federal courts have substantially clarified the analysis applied to franchise tying cases and have been reluctant to impose antitrust liability to problems with adequate remedies in contract law.

## **V. SELF-ENFORCEMENT AND OPTIMAL CONTRACT LAW**

Posner (2003) has recently noted that “economics fails to explain contract law” and that it “provides little normative guidance for reforming contract law.” That diagnosis is rather bleak for the law and economics of contracts. And, according to Posner, the economists are at least partly to blame. Economists have produced models that “focus on small aspects of contractual behavior or make optimal doctrine a function of variables that cannot be realistically observed, measured, or estimated.” The incomplete contracts literature fails to describe contracts themselves as its models predict contracts that are far more complex than those observed in the real world. On top of failing to explain real world contracts, the economics literature also fails to explain contract doctrine.

Posner is correct that much of what he describes as the incomplete contracts literature has failed to predict the content of actual contracts. Indeed, one recurring theme in this essay has been that the “conventional” incomplete contracting approach is unlikely to predict contract terms because it ignores the potential for these terms to facilitate self-enforcement. On the other hand, Posner essentially ignores the work of Klein, Williamson, Goldberg and others who have emphasized self-enforcement and reputation in understanding the role of contract terms. By way of contrast with the conventional approach, the self-enforcement framework has generated significant economic understanding of a

variety of contractual terms and complex distribution arrangements, including RPM, block booking, franchising, tying, slotting contracts, exclusive dealing, exclusive territories, and vertical integration. This literature provides a more optimistic view of the state and direction of the law and economics of contracts.

The key difference between the self-enforcement approach and the conventional approach is worth repeating yet again: the latter envisions the sole role of contract terms is to create the correct incentives for performance with respect to some contractually specified measure. Whereas, the self-enforcement framework implies that transactors will select contract terms to economize on their limited amounts of private enforcement capital and broaden the self-enforcing range to assure performance under the broadest possible range of likely ex post market conditions. Under the self-enforcement approach, contract terms not only are used to provide incentives to perform on some court-verifiable measure, but can also create rents that facilitate performance when coupled with the threat of termination. Because self-enforcement is an important feature of economic relationships, it is no surprise that the standard models have limited predictive value in explaining real-world contract terms. The standard approach continues, as a general matter, to proceed as if reputational enforcement is an “all or nothing” choice, rather than understanding the complementary nature of self- and court-enforcement mechanisms.

The self-enforcement framework Klein has developed over the last 30 years also has important implications for optimal contract law. In the extreme case, the standard approach would render the content of contract law largely irrelevant since parties could presumably contract around inefficient doctrine in most circumstances. But contract law is not irrelevant in the real world. The content of the law matters because contracting parties largely accept it as given and do not “opt-out” of all inefficient doctrine, thereby saving significant

resource costs by relying on what amounts to an exogenously imposed and impartial set of terms. And because the law matters, efficient contract doctrine sets rules that allow courts to identify terms parties would have adopted when contractual arrangements breakdown because some unspecified contingency occurs. In other words, efficient contract doctrine should set rules that allow the court to interpret an incomplete contract to reflect the terms the parties would have adopted ex ante had they contracted over those contingencies.

The self-enforcement framework might be especially useful for providing insights concerning the task of interpreting contracts. The economic approach to contracts poses a critical issue with respect to how judges should interpret intentionally incomplete contracts when the parties' intentions may deviate from the written terms. One proposal frequently raised in the law and economics literature is that judges should interpret contracts by filling in those terms parties would have adopted had they contracted over some contingency is obviously a difficult task. Despite widespread recognition of the difficulties associated with this sort of judicial contractual gap-filling, the self-enforcement analysis discussed throughout this essay suggests that law and economics scholars may have actually underestimated the complexity of this task. This is because the role of contract terms goes beyond shifting risks and creating direct incentives to preventing hold-up risk. Klein (1985) illustrates this point with the example of a simple long-term fixed-price contract. The conventional approach suggests that the sole role of the fixed price contract term is to insure against the risk of price volatility. However, the self-enforcement framework reveals that, in addition to allocating risk efficiently, fixed-price contracts can play an important role in reducing the probability of hold-up, and thereby encouraging specific investment and performance. Klein (1980) demonstrates that in order to economize on the parties' reputational capital, efficient self-enforcing

arrangements can often include seemingly “unfair” and “one-sided” contract terms that courts may have a tendency not to enforce.<sup>39</sup>

The self-enforcement framework implies that the primary role of contract law and interpretation in facilitating efficient economic exchange should, at least in principle, broaden the self-enforcing range. Contract law and interpretation doctrines can, in principle, be applied to minimize hold-up behavior by identifying attempts to hold-up a transacting party and preventing parties from using the court to facilitate a hold-up. Some have argued that contract doctrine has developed in a manner consistent with this view. For example, Muris (1981) argues that “when viewed through the lens of opportunism, many aspects of the law previously regarded as diverse in nature should be recognized as containing a common unifying principle” and that “judges can, and often do, act to lower important costs of transacting.”<sup>40</sup>

The application of contract interpretation to broaden the self-enforcing range has some appeal. The court could, in principle, operate as a substitute for the parties’ private reputational capital by going beyond literal interpretation methods to understand the context of the arrangement and parties’ intent. In this sense, contextual methods of interpretation have some theoretical advantage over more formal and literal methods, which involve enforcing terms only as they are written. However, the potential for judicial lowering of transactions costs comes with the risk that judges might apply contract law in a manner that reduces the self-enforcing range and, therefore, increases the costs to transacting parties and the brand-name capital required to avoid hold-up. It is a difficult problem for contract theorists and economists to understand the purpose of the

---

<sup>39</sup> See also Posner and Bebchuk (2006) for a similar analysis.

<sup>40</sup> On the law and economics of contract interpretation, see Goetz and Scott (1981), Schwartz (1992), Hadfield (1994), Ben-Sharar (1999), Schwartz and Scott (2003), Katz (2004), Posner (2005), Posner (1998), and Posner (2000).

terms parties adopt in the context of a complex contractual arrangement. It is difficult to imagine that the task would not be overwhelming for a judge without expertise in the particular industry or in the economics of contracts. Applying flexible interpretation methods, as predicted in Klein (1980), might lead courts to refuse to enforce non-standard or apparently one-sided contract terms that actually are important in encouraging specific investments and facilitating self-enforcement by shifting private capital.<sup>41</sup>

One weakness with the law and economics literature surrounding contract interpretation is that it has traditionally had little to do with contracts themselves rather than specific doctrines as they relate to parties' disputes about contracts. But how are courts to enforce the intent of the parties when the self-enforcement framework suggests that the "true" aims of the parties will often diverge, and possibly contradict, the contract's literal terms? The self-enforcement approach implies that if judicial efforts to align the true aims of the parties with written terms are ultimately erroneous, the result will be an increased strain on the parties' reputational capital and a more narrow self-enforcing range. The question of how willing courts are to enforce written contract terms, as well as how frequently judicial attempts to fill-gaps result in errors that frustrate the parties' intent, are ultimately empirical questions. However, the self-enforcement framework does suggest a more subtle, and perhaps more complex, approach to understanding contract terms and filling contractual gaps. In the absence of evidence that judges can competently carry out this difficult task, one might draw the tentative conclusion that the self-enforcement framework provides the basis for an economic argument in favor of formal contract interpretation methods.

---

<sup>41</sup> Hadfield (1994); Katz (2004); Posner (2005); Posner (2000); Posner (1998); Schwartz (1992).

## **VI. CONCLUSION**

Klein has made a number of seminal contributions to the law and economics of contracting. The cumulative impact of these contributions to our economic knowledge of contractual arrangements is substantial. Klein's work has improved our knowledge of franchise contracts, block booking, RPM, exclusive dealing, shelf space contracts, exclusive territories, and vertical integration. These contributions have not only improved our economic state of knowledge about contracts, but have significantly improved antitrust policy.

The sum of these individual contributions, however, understates the essence of Klein's contributions to law and economics. More important than any of these individual contributions is the analytical framework Klein developed to understand the role of contract terms in providing marginal incentives to perform against some court enforceable proxy, and also to facilitate self-enforcement by creating a premium stream that will be lost upon termination. This framework reconciled two stylized "facts" of contracting: (1) the ubiquity of incomplete contracts, and (2) the importance of private enforcement. Rather than assume away the critically important elements of contracting, Klein's work can be characterized as increasing our understanding of contractual arrangements in the real world. The same commitment to analysis of real world economics explains Klein's work in vertical restraints, and in particular, the rejection of the overly simplistic models retail competition in the conventional economics literature in favor of a more realistic approach that grapples with important institutional details. It is this more realistic approach, embracing rather than assuming away the complexities associated with the reality, which generated Klein's important economic insights about contractual relationships.

Contractual arrangements between sophisticated parties can be incredibly difficult to understand. This difficulty is further enhanced by understanding that

contractual arrangements, including vertical integration, do not merely provide incentives to perform, but also shift reputational capital between parties to facilitate performance. Increasing our economic knowledge of contracts and their terms in this complex environment requires significant attention to detail, knowledge of the law and relevant institutions, an understanding of the industry in which the contracts are adopted, as well as the economic incentives of the parties. Meaningful contributions do not come easily or without significant investment. Klein's pioneering work over the past 30 years has not only left us with a much greater understanding of contractual arrangements, but also a model for law and economics scholars and economists motivated to explain real world phenomenon, rather than merely producing blackboard insights.